



Collecte de parole pour l'étude des langues peu dotées ou en danger avec l'application mobile Lig-Aikuma

David Blachon, Elodie Gauthier, Laurent Besacier, Guy-Noël Kouarata,
Martine Adda-Decker, Annie Rialland

► To cite this version:

David Blachon, Elodie Gauthier, Laurent Besacier, Guy-Noël Kouarata, Martine Adda-Decker, et al.. Collecte de parole pour l'étude des langues peu dotées ou en danger avec l'application mobile Lig-Aikuma. Atelier Traitement Automatique des Langues Africaines (TALAF). JEP-TALN 2016, Jul 2016, Paris, France. hal-01350045

HAL Id: hal-01350045

<https://hal.science/hal-01350045>

Submitted on 29 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collecte de parole pour l'étude des langues peu dotées ou en danger avec l'application mobile LIG-AIKUMA

David Blachon¹, Elodie Gauthier¹, Laurent Besacier¹,
Guy-Noël Kouarata², Martine Adda-Decker², Annie Rialland²

(1) LIG Univ Grenoble Alpes, Grenoble, France

(2) LPP Univ Paris 3, Paris, France

{david.blachon, elodie.gauthier, laurent.besacier}@imag.fr
guy_kouarata@yahoo.com, madda@limsi.fr, annie.rialland@univ-paris3.fr

RÉSUMÉ

Nous rapportons dans cet article les travaux en cours portant sur la collecte de langues africaines peu dotées ou en danger. Une collecte de données a été menée à l'aide d'une version modifiée de l'application Android AIKUMA, initialement développée par Steven Bird et coll. (Bird *et al.*, 2014). Les modifications apportées suivent les spécifications du projet franco-allemand ANR/DFG BULB ¹ pour faciliter la collecte sur le terrain de corpus de parole parallèles. L'application résultante, appelée LIG-AIKUMA, a été testée avec succès sur plusieurs smartphones et tablettes et propose plusieurs modes de fonctionnement (enregistrement de parole, *respeaking* de parole, traduction et élicitation). Entre autres fonctionnalités, LIG-AIKUMA permet la génération et la manipulation avancée de fichiers de métadonnées ainsi que la prise en compte d'informations d'alignement entre phrases prononcées parallèles dans les modes de *respeaking* et de traduction. L'application a été utilisée aux cours de campagnes de collecte sur le terrain, au Congo-Brazzaville, permettant l'acquisition de 80 heures de parole. La conception de l'application et l'illustration de son usage dans deux campagnes de collecte sont décrites plus en détail dans cet article.

ABSTRACT

Parallel speech collection for under-resourced language studies using the LIG-AIKUMA mobile device app

This paper reports on our ongoing efforts to collect speech data in under-resourced or endangered languages of Africa. Data collection is carried out using an improved version of the Android application AIKUMA developed by Steven Bird and colleagues (Bird *et al.*, 2014). Features were added to the app in order to facilitate the collection of parallel speech data in line with the requirements of the French-German ANR/DFG BULB project (Breaking the Unwritten Language Barrier). The resulting app, called LIG-AIKUMA, runs on various mobile phones and tablets and proposes a range of different speech collection modes (recording, *respeaking*, translation and elicitation). LIG-AIKUMA's improved features include a smart generation and handling of speaker metadata as well as *respeaking* and parallel speech data manipulation. It was used for field data collections in Congo-Brazzaville resulting in a total of over 80 hours of speech. Design issues of the mobile app as well as the use of LIG-AIKUMA during two recording campaigns, are further described in this paper.

MOTS-CLÉS : Collecte de données, application mobile, documentation de langues, corpus oraux.

KEYWORDS: Speech collection, mobile application, language documentation, spoken corpora.

1. <http://www.bulb-project.org>

1 Introduction

L'essor considérable des smartphones et autres appareils mobiles offre de nouvelles opportunités aux linguistes de terrain et aux chercheurs dans le domaine de la documentation des langues. Il est désormais possible d'envisager des campagnes pour la documentation des langues en danger, impliquant de nombreux locuteurs et permettant l'acquisition de nombreuses ressources pour étudier divers phénomènes linguistiques (analyse de la phonologie, tonologie, morphologie, lexicologie et lexicographie, syntaxe, pragmatique). Les enregistrements de terrain pourraient aussi permettre la collecte de ressources ethno-linguistiques qui ont une importance particulière pour la documentation de traditions et des modes de vie. Cependant, des collectes de données à grande échelle nécessitent une organisation optimale des ressources pour faciliter l'accès aux données selon des conventions strictes de nommage de fichiers et de méta-données.

Dans la suite de l'article, nous décrivons la réalisation d'une application mobile pour l'enregistrement efficace de ressources linguistiques dans l'environnement naturel des locuteurs. L'application génère de manière automatique des fichiers de parole et des fichiers de méta-données associés. Nous décrivons également notre retour d'expérience dans la réalisation récente de deux campagnes de collecte.

1.1 Le projet BULB

Le projet BULB (Breaking the Unwritten Language Barrier) s'intéresse au problème de la documentation de langues non-écrites en s'appuyant sur l'usage de techniques de reconnaissance automatique de parole et de traduction automatique. Le projet s'appuie sur une étroite collaboration franco-allemande composée de linguistes et de chercheurs en informatique. L'objectif du projet est de concevoir et d'adapter des méthodes issues du traitement automatique de la langue (TAL) pour aider à l'analyse linguistique de langues ou dialectes, peu ou non écrits. Dans le projet, la famille des langues bantoues est visée en particulier. Cette famille inclut environ 500 langues (les chiffres varient de 440 (Guthrie, 1971) à plus de 660 (Mann & Dalby, 1987; Maho, 2003) suivant les auteurs), dont la plupart sont essentiellement orales et n'ont pas encore été étudiées. À cette fin, trois langues partiellement étudiées et disposant de peu de ressources ont été choisies : le basaa (A43), le myene (B11) et le mbochi (C25). Dans cet article, nous rapportons notre expérience de collecte de données en mbochi avec LIG-AIKUMA.

1.2 Le concept de *respeaking*

Bird et coll. (Bird, 2010) ont exploité le concept de *respeaking*, initialement introduit par Woodbury (Woodbury, 2003), afin de pouvoir collecter des données langagières dans des villages isolés. Leur idée était de récupérer des données culturelles et linguistiques représentatives des langues en danger, même en étant éloigné de toute forme de ressource numérique. Ainsi, des locuteurs d'une langue papoue locale ont été entraînés à la méthode dite de *respeaking*. Cette méthode nécessite d'écouter un enregistrement sonore original et de répéter précautionneusement son contenu. Un second enregistrement en découle, plus facile à traiter ultérieurement, que ce soit par un linguiste ou par un système automatique. Les motivations pour le *respeaking* sont multiples : la parole de

l'enregistrement original peut être trop rapide, le niveau d'enregistrement trop faible ou du bruit peut dégrader le contenu. De plus, une tâche comme l'enregistrement de contes ou histoires nécessite souvent la participation de personnes âgées qui peuvent avoir une voix faible ou une prononciation altérée (dents manquantes, etc.), ce qui peut compromettre la qualité de l'enregistrement (Bird, 2014).

1.3 Résumé de l'article

L'article est organisé de la manière suivante. La section 2 décrit l'application initiale qui a servi de base au développement de LIG-AIKUMA, elle-même présentée dans la section 3. La section 4 résume les premières collectes de données menées au Congo-Brazzaville. Enfin, la section 5 conclut et propose quelques perspectives.

2 Les origines d'AIKUMA

2.1 L'application initiale AIKUMA et ses motivations

Bird et coll. (Bird, 2014) ont développé l'application initiale AIKUMA afin de permettre l'exploitation sur le long terme des ressources recueillies. Selon les auteurs, l'application est conçue pour un *philologue du futur* (Bird et al., 2014) : elle peut collecter assez de parole documentée pour permettre un traitement postérieur par le linguiste. En effet, les auteurs observent qu'en règle générale, les projets de documentation de langues manquent de ressources, en particulier d'experts humains pour effectuer les traitements postérieurs nécessaires sur les données. En conséquence, lesdits traitements peuvent avoir lieu très longtemps après la collecte des données, ce qui peut mener à des situations délicates : par exemple, il est possible qu'une langue en danger ait disparu entre temps et qu'il n'y ait donc plus aucun locuteur de la langue pour réaliser lesdits traitements. Ces raisons ont mené les auteurs à étendre le concept de *respeaking* à une traduction orale des énoncés. Cette traduction est effectuée suivant le même principe que le *respeaking* et consiste à répéter un enregistrement dans une langue différente de la langue de l'enregistrement original.

AIKUMA est une application conçue pour l'environnement mobile Android. Elle propose les fonctionnalités suivantes : enregistrement de parole, *respeaking* et traduction orale d'un enregistrement sonore. Ces fonctionnalités sont décrites dans la section suivante.

2.2 Fonctionnalités de l'application initiale

Dans AIKUMA, les tâches de *respeaking* de parole et de traduction orale fonctionnent de la même manière : le locuteur écoute les segments d'un fichier original, puis prononce ce qu'il a entendu, dans la même langue (tâche de *respeaking*) ou dans une autre (tâche de traduction). Les segments de parole produits sont concaténés les uns aux autres pour former un seul fichier de *respeaking* ou de traduction orale. Une fois concaténés dans un fichier, il peut être difficile de distinguer les segments, c'est pourquoi un fichier textuel contenant les informations d'alignement de tous les segments des deux fichiers sonores est aussi produit. La figure 1 illustre un tel alignement de segments.

L'application a été pensée pour une utilisation large et une analyse future. Ainsi, l'interface d'utilisation est composée d'icônes sans mention de texte, pour permettre l'utilisation par un grand nombre

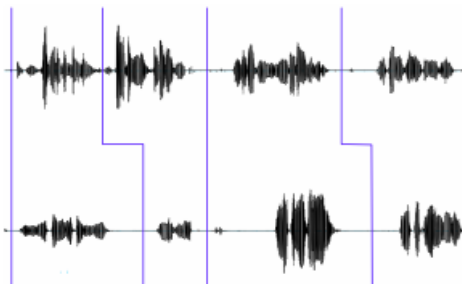


FIGURE 1 – Illustration de l’alignement des segments de deux fichiers sonores (Bird *et al.*, 2014).

de personnes. L’application propose aussi une fonctionnalité de synchronisation et de partage sans fil qui permet à plusieurs appareils de partager des fichiers audio. Celle-ci est particulièrement utile pour étendre les tâches de *respeaking* et de traduction à de larges groupes de personnes. En plus des fichiers sonores, l’application permet l’enregistrement de métadonnées pour la documentation des fichiers, qui incluent notamment des informations sur le locuteur, la langue et le lieu d’enregistrement.

Les auteurs indiquent avoir utilisé AIKUMA dans des communautés indigènes reculées (notamment au Brésil et au Népal (Bird, 2014)).

3 L’évolution d’AIKUMA vers LIG-AIKUMA

3.1 Motivations et spécifications

Avec un outil comme AIKUMA, les objectifs du projet BULB ont donné lieu à des cas d’utilisation spécifiques : l’enregistrement sonore libre ; le *respeaking* et la traduction orale ; et enfin, l’éllicitation de parole par l’affichage de textes, d’images ou de vidéos.

Pour permettre la réalisation des cas d’utilisation précédents, certaines adaptations à l’application existante ont été proposées. D’abord, l’interface utilisateur a été modifiée de manière à identifier et sélectionner plus facilement une tâche. Également, toujours suivant l’objectif de rendre l’application rapide et facile à utiliser, une fonctionnalité d’enregistrement et de chargement des métadonnées du dernier enregistrement a été proposée. Ainsi, lorsqu’un utilisateur souhaite effectuer un nouvel enregistrement avec le même locuteur, le formulaire de métadonnées est pré-rempli. Les sections ci-après décrivent les fonctionnalités les plus importantes réalisées dans LIG-AIKUMA.

3.2 Les modes d’utilisation

Les fonctionnalités essentielles de l’application AIKUMA initiale ont été conservées pour l’enregistrement, le *respeaking* de parole et la traduction, ainsi que la sauvegarde des métadonnées sur le locuteur. En outre, certaines parties de l’interface ont été réutilisées.

Par ailleurs, de nouveaux développements ont mis l’accent sur la mise en place de 4 modes, dédiés à des tâches spécifiques d’enregistrement de la parole. L’écran d’accueil est illustré sur la figure 2 (à

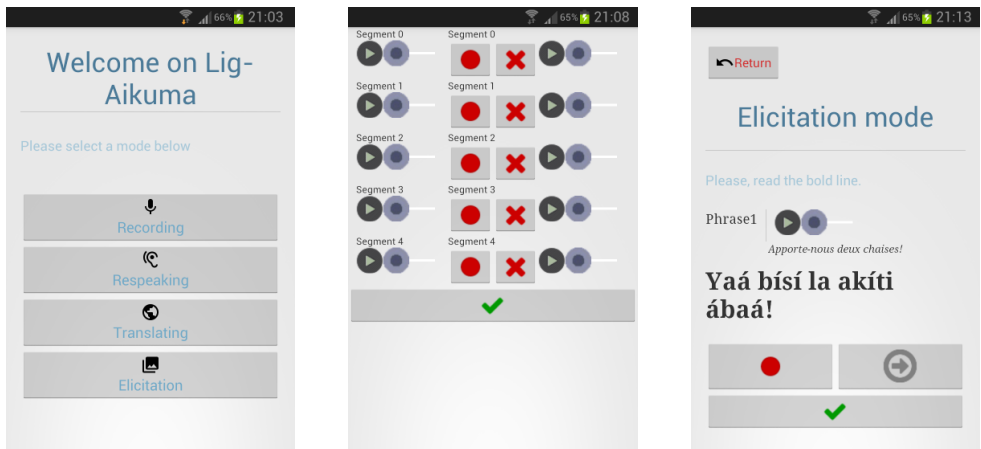
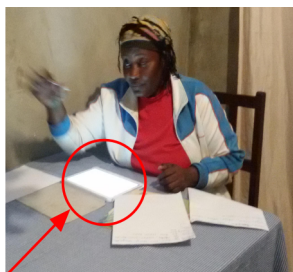


FIGURE 2 – Captures d’écran de l’application LIG-AIKUMA : de gauche à droite, 1) la page d’accueil ; 2) la vue du résumé des segments après une tâche de *respeaking* ou de traduction : l’utilisateur peut écouter ou reproduire tous les segments ; 3) le mode dit "d’éllicitation" qui incite à la production de parole après affichage de texte.

gauche). Comme on peut le voir, les quatre modes suivants sont identifiés :

- Enregistrement libre de parole spontanée (*recording* en anglais),
- *Respeaking* d’un enregistrement : cet enregistrement peut provenir d’un fichier enregistré avec l’application ou d’un fichier sonore externe ; la nouvelle fonctionnalité de *respeaking* de parole permet désormais d’écouter (optionnellement) le dernier segment d’enregistrement afin de le vérifier et de le reproduire si nécessaire, avant de passer au segment suivant. De plus, une fois que la tâche est terminée, une vue résume l’ensemble des nouveaux segments et leur version originale ; cette vue permet aussi (le cas échéant) d’écouter et de réenregistrer un segment avant de terminer la session. Sur la figure 2 (au milieu), on peut voir que les segments originaux sont alignés avec ceux qui sont reproduits. Les deux types de segments peuvent être lus mais seuls les segments reproduits peuvent être modifiés, ce qui permet une double vérification et la correction d’erreurs,
- Traduction d’un enregistrement (*translating* en anglais) : cette tâche est identique à celle du *respeaking*, à l’exception de la langue de la traduction, qui doit être différente de celle du fichier original,
- Elicitation à partir d’un texte : l’utilisateur charge un fichier texte dans l’application, puis lit la phrase, la prononce, écoute l’enregistrement et passe à la suivante. Ce mode a été spécifiquement requis pour la collecte de données qui a eu lieu au Congo-Brazaville pendant l’été 2015 (décrite en section 4). Les perspectives de développement de ce mode prévoient d’intégrer des images et des vidéos pour l’éllicitation. La figure 2 (droite) illustre ce mode.



Aikuma on Android tablet



Aikuma on Android tablet

FIGURE 3 – Exemples d'utilisation de l'application LIG-AIKUMA sur des tablettes Android au cours d'une campagne de collecte : à gauche, la prononciation de conjugaisons de verbes par une locutrice native du mbochi et à droite, des conversations spontanées entre plusieurs locuteurs.

3.3 Métadonnées et noms de fichiers

Dans chaque mode, un fichier de métadonnées est enregistré avec le fichier sonore. Les métadonnées sont renseignées dans un formulaire avant la réalisation d'une nouvelle tâche. Contrairement à la version initiale de AIKUMA qui proposait quatre vues différentes pour cela, le formulaire rassemble toutes les informations au sein d'une interface unique.

Par ailleurs, les métadonnées ont été enrichies avec de nouvelles informations sur les langues (langue de l'enregistrement, langue maternelle du locuteur et autres langues parlées) et sur le locuteur (nom, âge, sexe, région d'origine). D'autre part, afin de gagner du temps, une fonctionnalité permet de sauvegarder les dernières informations renseignées dans une session et de les charger dans le formulaire lors de l'utilisation suivante.

Les fichiers sont actuellement nommés suivant le format : *DATE-HEURE_NOM-APPAREIL_LANGUE*. Par exemple, *2015_07_22_17_00_00_Samsung_tablette_fra* est le nom d'un enregistrement effectué le 22 juillet 2015 à 17h (17h00m00), en langue française, sur une tablette Samsung.

3.4 Autres améliorations

L'interface a été adaptée pour les grands écrans des tablettes (10 pouces), de sorte que l'application fonctionne aussi bien sur smartphones que sur tablettes Android. Outre les spécifications présentées, basées sur de multiples discussions avec des linguistes de terrain, cette nouvelle version a été développée en environ 3 hommes / mois et a généré plus de 5000 lignes de code. Le nouveau code est sauvegardé sur la forge du LIG² et est accessible à la demande, pour l'utilisation ou le développement. Suivant la licence originale de l'application AIKUMA, LIG-AIKUMA est soumise à la licence de diffusion GNU Affero General Public License³.

L'application LIG-AIKUMA a été installée avec succès sur plusieurs appareils, incluant notamment un Samsung Galaxy SIII, un Google Nexus 6, un HTC Desire 820 et des tablettes Galaxy Tab 4.

2. <https://forge.imag.fr/projects/lig-aikuma/>

3. <https://www.gnu.org/licenses/agpl-3.0.en.html>

3.5 Téléchargement de LIG-AIKUMA

Les utilisateurs qui souhaitent seulement utiliser l'application, sans accès au code, peuvent télécharger le fichier directement à partir du lien de la forge⁴. Pour ce faire, il est nécessaire de taper l'adresse du lien précédent dans un navigateur Web ; une fenêtre apparaît pour demander la confirmation de l'installation (au préalable, l'utilisateur doit avoir autorisé l'installation à partir d'autres sources que Play Store).

4 Première campagne d'acquisition

4.1 Objectifs et préparation des ressources

Une campagne de collecte de données sur le terrain nécessite une préparation particulière. Dans le projet BULB, il était important de considérer à la fois les exigences linguistiques et technologiques, puisque l'objectif général est de développer des outils de TAL pour le linguiste de terrain. Avant de démarrer la campagne, un cahier des charges résumant les deux types de besoins a été établi.

Au-delà des questions éthiques habituelles et des formulaires de consentement, les spécifications ont porté sur la collecte de « données linguistiquement denses », par exemple, la prononciation de conjugaisons de verbes (Kouarata, 2014) ou de petites phrases, dans le but de recueillir des données pertinentes pour la description de la langue orale, et suivant les méthodes habituelles de collecte sur le terrain. La figure 3 illustre l'utilisation de LIG-AIKUMA sur une tablette pour enregistrer des conjugaisons en mbochi (à gauche) et des conversations spontanées entre locuteurs (à droite). L'objectif initial était de collecter de grandes quantités de données (plusieurs dizaines d'heures) de dizaines de locuteurs dans différents styles de parole. Tous les documents écrits possibles, incluant un dictionnaire mbochi-français de 3200 entrées (Beapami *et al.*, 2000), des contes traditionnels et des textes bibliques en mbochi ont été regroupés et ont servi à la collecte. En outre, 1200 des 6000 phrases de référence de la documentation orale du langage (Bouquiaux & Thomas, 1976) ont été traduites et écrites en mbochi par l'un des auteurs (Guy-Noël Kouarata). Ces sources ont été adaptées pour pouvoir être exploitées dans le mode d'éllicitation de LIG-AIKUMA. Enfin, un autre objectif de la campagne sur le terrain était de collecter photos ou vidéos correspondant à des coutumes culturelles traditionnelles. Ces captures pourront être utilisées comme des ressources supplémentaires pour l'éllicitation dans des collectes ultérieures. Elles peuvent aussi servir à illustrer des ressources linguistiques, telles que les dictionnaires en ligne et du matériel éducatif, ce qui constitue un retour de valeur pour la communauté de locuteurs qui a accepté de contribuer à l'étude.

4.2 Données collectées

En pratique, deux campagnes d'un mois au Congo Brazzaville ont été effectuées par l'un des auteurs (Guy-Noël Kouarata) qui est un locuteur natif du mbochi. La campagne d'enregistrement des données a été organisée comme suit : quatre tablettes *Galaxy Tab 4* ont été prêtées à quatre personnes de contact qui en étaient responsables et dont le rôle était de recruter des locuteurs volontaires bénévoles parmi leurs proches, amis et connaissances. Il était également demandé aux personnes de contact de contribuer à la tâche de *respeaking*. Tous les locuteurs ont été payés et les personnes de contact gardaient les tablettes à la fin du processus de collecte, comme récompense supplémentaire.

4. <https://forge.imag.fr/frs/download.php/706/MainActivity.apk>

Jusqu'à maintenant, le projet BULB a permis d'enregistrer 48 heures de données de parole en langue mbochi avec LIG-AIKUMA au Congo-Brazzaville. Le corpus est composé de 33 heures de parole spontanée (surtout des débats et des histoires), 9,5 heures de parole contrôlée (incluant des conjugaisons), 2 heures de phrases lues (recueillies à l'aide du mode d'élicitation) et 3,5 heures de lecture. Une autre collecte de prononciation de 1000 phrases a été réalisée, ainsi que l'enregistrement de traductions d'énoncés en français.

5 Conclusion

Cet article a présenté LIG-AIKUMA, une application de collecte de parole pour la documentation de langues peu dotées ou en danger. Grâce à celle-ci, une première collecte de ressources linguistiques en mbochi a été menée et de nouveaux développements et fonctionnalités sont prévus dans un avenir proche. L'un des objectifs du projet BULB est aussi l'extraction automatique de connaissances à partir des enregistrements sonores parallèles obtenus par le *respeaking* et la traduction. Cela pourrait être utilisé, par exemple, pour apprendre de manière automatique les unités de mots (ainsi que leur prononciation) dans une langue inconnue (et non écrite) avec très peu de supervision.

Remerciements

Ce travail a été réalisé dans le cadre du projet franco-allemand ANR-DFG BULB (ANR-14-CE35-002) et soutenu financièrement par le programme français d'Investissements d'Avenir - Labex EFL (ANR-10-LABX-0083).

Références

- BEAPAMI R. B., CHATFIELD R., KOUARATA G. & EMBENGUE-WALDSCHMIDT A. (2000). *Dictionnaire Mbochi-Français*. Congo (Brazzaville) : SIL-Congo Publishers.
- BIRD S. (2010). A scalable method for preserving oral literature from small languages. In *Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries*.
- BIRD S. (2014). : COLING.
- BIRD S., HANKE F. R., ADAMS O. & LEE H. (2014). Aikuma : A mobile app for collaborative language documentation. *ACL 2014*, p.1.
- BOUQUIAUX T. & THOMAS J. (1976). *Enquête et description des langues à tradition orale*. Paris : SELAF.
- GUTHRIE (1971). *Comparative Bantu, Volume 2*. Farnborough : Gregg Press.
- KOUARATA G. (2014). *Variations de formes dans la langue Mbochi (Bantu C25)*. Thèse de doctorat, Université Lumière Lyon 2, Lyon.
- MAHO J. (2003). *A classification of the Bantu languages : an update of Guthrie's referential system*, In D. NURSE & G. PHILIPPSON, Eds., *The Bantu Languages*, p. 639–651. Routledge.
- MANN M. & DALBY D. (1987). *A thesaurus of African languages*. London : Hans Zell Publishers.
- WOODBURY A. C. (2003). *Defining documentary linguistics*, In P. K. AUSTIN, Ed., *Language Documentation and Description*, volume 1, p. 35–51. Language Documentation and Description, SOAS.